# NAVAL POSTGRADUATE SCHOOL

**MONTEREY, CALIFORNIA**

# THESIS

**DATA ANALYSIS OF CRITERIA GOVERNING SELECTION OF ACTIVE GUARD/RESERVE COLONEL**

by

Joseph A. Snel

September 2014

Thesis Advisor: Rachel Silvestrini
Second Reader: Lyn Whitaker

THIS PAGE INTENTIONALLY LEFT BLANK

| **REPORT DOCUMENTATION PAGE** | | | *Form Approved OMB No. 0704–0188* |
|---|---|---|---|
| Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704–0188) Washington DC 20503. | | | |
| **1. AGENCY USE ONLY *(Leave blank)*** | **2. REPORT DATE**<br>September 2014 | **3. REPORT TYPE AND DATES COVERED**<br>Master's Thesis | |
| **4. TITLE AND SUBTITLE**<br>DATA ANALYSIS OF CRITERIA GOVERNING SELECTION OF ACTIVE GUARD/RESERVE COLONEL | | | **5. FUNDING NUMBERS** |
| **6. AUTHOR(S)** Joseph A. Snel | | | |
| **7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**<br>Naval Postgraduate School<br>Monterey, CA 93943–5000 | | | **8. PERFORMING ORGANIZATION REPORT NUMBER** |
| **9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)**<br>N/A | | | **10. SPONSORING/MONITORING AGENCY REPORT NUMBER** |
| **11. SUPPLEMENTARY NOTES** The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____. | | | |
| **12a. DISTRIBUTION / AVAILABILITY STATEMENT**<br>Approved for public release; distribution is unlimited | | | **12b. DISTRIBUTION CODE** |
| **13. ABSTRACT (maximum 200 words)** | | | |
| **14. SUBJECT TERMS** Data Analysis, Logistic Regression, Active Guard/Reserve, Promotion Selection | | | **15. NUMBER OF PAGES**<br>71 |
| | | | **16. PRICE CODE** |
| **17. SECURITY CLASSIFICATION OF REPORT**<br>Unclassified | **18. SECURITY CLASSIFICATION OF THIS PAGE**<br>Unclassified | **19. SECURITY CLASSIFICATION OF ABSTRACT**<br>Unclassified | **20. LIMITATION OF ABSTRACT**<br>UU |


This thesis provides data analysis on the selection process of the FY 2009–2011 Army Active Guard/Reserve (AGR) colonel selection boards. In this analytic study, logistic regression is used to study what variables influence colonel selection. The focus of this study is to aid Army senior leaders in the mentoring and development of future senior leaders by identifying criteria key to the selection of Army AGR colonels. A data set is compiled from 1144 individual promotion packets submitted across three selection boards. The 1144 packets correspond to 684 individuals. The findings suggest one's zone of consideration, age, longest deployment, senior service college completion, possession of a master's degree, battalion command, number of ratings as a lieutenant colonel, and the total percentage above center of mass ratings have a significant influence on selection.


NSN 7540–01–280–5500

Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

**Approved for public release; distribution is unlimited**

# DATA ANALYSIS OF CRITERIA GOVERNING SELECTION OF ACTIVE GUARD/RESERVE COLONEL

Joseph A. Snel
Lieutenant Colonel, United States Army Reserve
B.S., Norwich University, 1993

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN OPERATIONS RESEARCH**

from the

**NAVAL POSTGRADUATE SCHOOL**
**September 2014**

Author: Joseph A. Snel

Approved by: Rachel Silvestrini
Thesis Advisor

Lyn Whitaker
Second Reader

Robert F. Dell
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

# ABSTRACT

This thesis provides data analysis on the selection process of the FY 2009–2011 Army Active Guard/Reserve (AGR) colonel selection boards. In this analytic study, logistic regression is used to study what variables influence colonel selection. The focus of this study is to aid Army senior leaders in the mentoring and development of future senior leaders by identifying criteria key to the selection of Army AGR colonels. A data set is compiled from 1144 individual promotion packets submitted across three selection boards. The 1144 packets correspond to 684 individuals. The findings suggest one's zone of consideration, age, longest deployment, senior service college completion, possession of a master's degree, battalion command, number of ratings as a lieutenant colonel, and the total percentage above center of mass ratings have a significant influence on selection.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

## LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

## LIST OF TABLES

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| %CW | Percent Total Conventional Wisdom |
| ACOM | above center of mass |
| AGR | Active Guard/Reserve |
| AWC | Army War College |
| AWC DL | Army War College Distance-Learning |
| CNW | College of Naval Warfare |
| CW1 | conventional wisdom 1 |
| CW2 | conventional wisdom 2 |
| CW3 | conventional wisdom 3 |
| CW4 | conventional wisdom 4 |
| CW5 | conventional wisdom 5 |
| DA PAM | Department of the Army Pamphlet |
| DCSPER | Deputy Chief of Staff for Personnel |
| DOD | Department of Defense |
| FY | fiscal year |
| FTS | Full Time Support |
| ICAF | Industrial College of the Armed Forces |
| IRB | Institutional Review Board |
| JAWS | Joint Advanced Warfighter Course |
| LTC | lieutenant colonel |
| MSM | Meritorious Service Medal |
| MT | military technician |
| NWC | National War College |
| OCAR | Office of the Chief, Army Reserve |
| OER | Officer Evaluation Report |
| PA&E | Program Analysis & Evaluation |
| SSC | Senior Service College |
| SSC F | Senior Service College Fellowship |
| Std Error | standard error |
| ToS | time on station |

THIS PAGE INTENTIONALLY LEFT BLANK

# EXECUTIVE SUMMARY

As the country faces the historically cyclic, post-war draw-down in military strength coupled with a reduction in budget, it is critical for leaders to possess an efficient means to facilitate the decision-making process in the selection of its future leaders. Draw-downs lend to an exodus of well-trained, experienced future senior leaders within the military ranks. To combat this, mentoring is crucial and providing the right conventional wisdom is necessary in leader development.

This thesis provides data analysis governing the selection process of the FY 2009–2011 Army Active Guard/Reserve (AGR) colonel selection boards. In this analytic study, logistic regression is used to examine what variables, if any, influence colonel selection. The focus of this study is to aid Army senior leaders in the mentoring and development of future senior leaders by means of identifying criteria key to the selection process for Army AGR colonels.

The Directorate of Program Analysis and Evaluation (PA&E), Office of the Chief, Army Reserve (OCAR) conducted a study in July of 2012, on the criteria necessary for selection of AGR lieutenant colonels to colonel. Information regarding 1144 promotion packets presented during the FY 2009–2011 AGR Colonel Boards were compiled to describe the characteristics of officers selected for promotion and determine the relevant factors influencing selection.

The data, provided by PA&E, contains 59 fields which are reduced to 33 fields for this study. The 1144 packets correspond to 684 individuals according to the identification number included in the data. The 684 individuals correspond to 321 one-time submissions, 266 two-time board submissions, and 97 three-time board submissions. In total, 170 packets were selected for promotion to colonel; representing 25% of all packets submitted as selected over the three-year period. This thesis supports the study of the 2009–2011 AGR Colonel Board analysis by providing an additional logistic regression study.

Logistic regression is a powerful data analysis tool for modeling outcomes of a Bernoulli random variable. Thus, logistic regression is an effective tool for modeling promotion.

The three measures of effectiveness used in this study focus on the logistic regression prediction percentages associated with being Correct, False-Positive and False-Negative. The classification of False-Positive is measured based upon a models predicted outcome of 1% or less. The classification of False-Negative is measured based upon a models predicted outcome of 15% or less. The intersection of the False-Positive and False-Negative outcomes is used to identify the ideal threshold of the confusion matrix for each fitted model. The correct prediction percentage is used in comparison between the fitted model outcomes.

The findings suggest one's zone of consideration, age, longest deployment, senior service college completion, possession of a master's degree, battalion command, number of ratings as a lieutenant colonel, and the total percentage above center of mass ratings have an influence on selection. The logistic regression models have an accuracy of prediction ranging from 83.04% to 89.33% with a False-Positive classification rate of 0.58% to 4.53%. Of the variables included in the logistic regressions, four are from a collection of "Conventional Wisdom" variables that capture what was perceived to be the most needed traits to be selected for promotion to colonel. When used alone, the conventional wisdom variables produce a logistic regression model with 82% accuracy.

# ACKNOWLEDGMENTS

I am ever so thankful for all those who supported the creation of this document. First and foremost, I am thankful for the God-given gift of analysis. Secondly, to my family, thank you for the sacrifice of my absence in the completion of my degree. I wish to thank the faculty and staff of the Naval Postgraduate School for their wisdom and patience in teaching this "old-dog-new-tricks." Special thanks go to Professors Rachel Silvestrini and Lyn Whitaker for guiding me through the thesis process. I thank LTCs Weko and Pontius, whose initial work is the foundation for this study. Finally, I would like to recognize my fellow classmates for the long hours and dedication to "leaving no one behind."

THIS PAGE INTENTIONALLY LEFT BLANK

# I. INTRODUCTION

## A. PURPOSE

This thesis provides data analysis governing the selection process of the FY 2009–2011 Army Active Guard/Reserve (AGR) colonel selection boards. In this analytic study, logistic regression is used to examine what variables, if any, influence colonel selection. The focus of this study is to aid Army senior leaders in the mentoring and development of future senior leaders by means of identifying criteria key to the selection process for Army AGR colonels.

## B. BACKGROUND

The AGR program was originally designed to support unit level activities and provide administrative support to the unit and headquarters levels. This support came in the form of "organizing, administering, recruiting, instructing, or training the reserve forces" (England, 1984, p. 11). At the time, a career in the AGR program was not part of the plan, thus it was uncommon to find senior ranking AGR members, especially colonels. This all changed upon the conversion of the Military Technician program into the newly established AGR program and was later followed by a demand for the increased roles and responsibilities of the AGR.

The Army Reserve Military Technician (MT) program is the forerunner to the AGR program. Established in 1950 (U.S. General Accounting Office, 1982), the program was instituted to provide a steady-state of operations for Reserve units during non-training periods. The positions were filled by civilians with no associated military obligations. Over the course of the next 20 years, and two official memorandums of understanding, the program evolved into the framework for today's civilians who work directly for Reserve units. The United States General Accounting Office highlighted the newly developed dual status program in its 1982 report to Congress stating the MT's role is to "maintain operations and training status of Reserve units." And "as a condition of employment, to participate in military training drills one weekend a month and about 2

weeks annually as military members—drilling reservists—of their units…are placed on active duty upon mobilization, and they should deploy with their units as military personnel" (U.S. General Accounting Office, 1982, p. 2).

The report also identified a discrepancy in end-strength accountability. The MT's were being counted in their civilian capacity as well as when they were on drilling status. This discrepancy was in non-compliance with the directives established by Public Law 93–365 (Department of Defense (DOD) Appropriation Authorization Act of 1975). Additionally, DOD Directive 1100.4, dated August 1954, outlined the position requirements of civilian personnel which later were determined as an incompatibility with the needs of the Army Reserve. Reports conducted by manpower commissions and several appropriations committees determined the negative impacts to the Army Reserve and the military as a whole, if a military technician were retained as opposed to conversion to AGR positions.[1]

As a result of the congressional concerns governing reserve recruitment; reserve readiness; problems relative to MTs; and the proper classification of military personnel, the AGR program came into existence. The authorization for this new military personnel classification is found under the DOD Authorization Act, 1980, Pub. L. No. 96–107, 0 401(b), 93 Stat. 807 (England, 1984). In response to congressional concern regarding reserve forces readiness, the Office of the Secretary of Defense directed an increase in Full-Time Support (FTS), mostly comprised of MTs, from its 5,800 end-strength. The strength, as of FY 2012, is 2.8 times that of the 5,800 total in 1979. This increase in strength is depicted in Figure 1, showing the Army Reserve end-strength Post-World War II to the present.

[1] Further details relative to the conversion of military technicians to the AGR program can be found via the report by the U.S. General Accounting Office

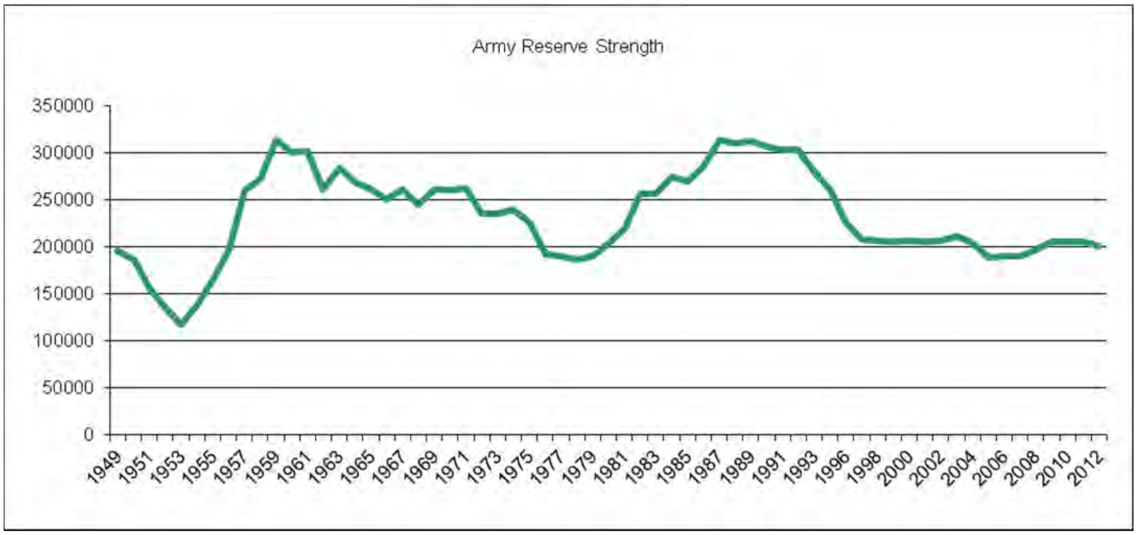


Figure 1. Army Reserve Strength by U.S. Army Reserve Command Headquarters (from LTC David Cloft, n.d.)

> In 1983, the Deputy Chief of Staff for Personnel (DCSPER) of the Army directed a study group to develop a methodology for assessing the increased need for AGR personnel and develop a 'feasible management framework' for the AGR program. This management framework must include the total life cycle of AGR members from accessioning to separation or retirement. (England, 1984, p. 13)

The introduction of a career AGR along with the opportunities for AGR's to hold competitive positions, as those of commanders, outside of the originally mandated administrative and support roles, leads to the organization of career development paths running parallel to both Reserve and Active Duty career progression, since an AGR Soldier is counted against the Reserve Force end-strength while in an Active Duty status. Figure 2, outlines the career path of a Reserve Officer, specifically that of an Engineer, as set for FY 2010. Similar career paths, based on branch affiliation, were utilized by those individuals submitting packets for promotion selection to colonel and whose packets and promotion results are examined in this thesis.

The Active and Reserve Components of the Army do not share quite the same career paths, according to the Commissioned Officer Professional Development and Career Management, Department of the Army Pamphlet 600–3, mostly due to actual time/experience spent in service and the difference in available duty positions. The AGR

program, although not a separate component of the Army, is a hybrid of the two components and requires a development process in and of its own.

An officer can now remain in the AGR program to retirement and compete for duty positions to broaden their careers into areas with greater rank, influence, and visibility; as that of a colonel. Criteria for selection to colonel in the AGR program should be identified and assessed against a comparison of both the Active and Reserve selection criteria standards. It is vital that the Army maintains a viable developmental program to ensure the proper mentoring of its leadership as the AGR program increases its end-strength quotas into the influential and policy making ranks of colonel.

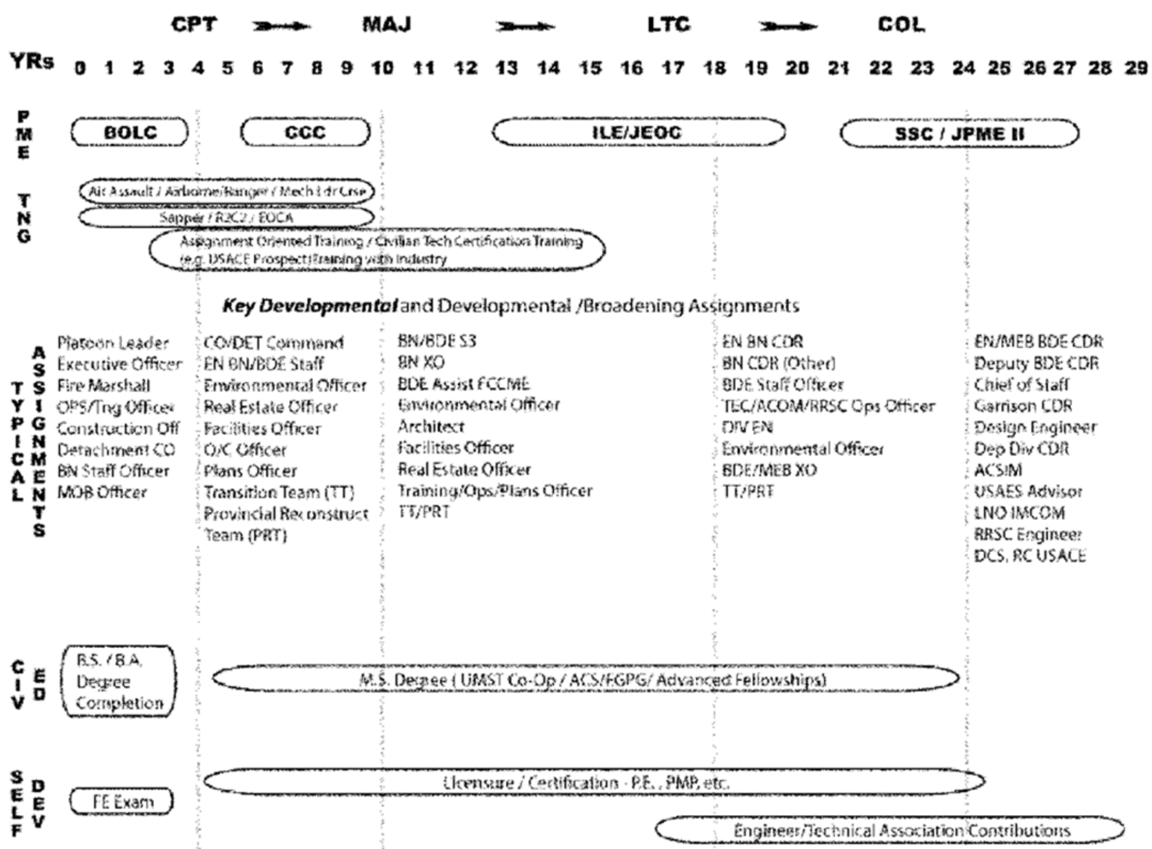


Figure 2. The Reserve Component Engineer Officer Development Model, (from DA PAM 600–3 Figure 14–4)

The Directorate of Program Analysis and Evaluation (PA&E), Office of the Chief, Army Reserve (OCAR) conducted a study in July of 2012, on the criteria

necessary for selection of AGR lieutenant colonels to colonel. Information regarding 1144 files presented during the FY 2009–2011 AGR Colonel Boards were compiled to describe the characteristics of officers selected for promotion and determine the relevant factors influencing selection. Results of the study generated interest in further analysis. This thesis supports the study of the 2009–2011 AGR Colonel Board analysis by providing an additional logistic regression study.

## C. SUMMARY

As the country faces the historically cyclic, post-war draw-down in military strength coupled with a reduction in budget, it is critical for leaders to possess an efficient means to facilitate the decision-making process in the selection of its future leaders. Draw-downs lend to an exodus of well-trained, experienced future senior leaders within the military ranks.[2] To combat this, mentoring is crucial and providing the right direction is necessary in leader development. In addition to determining whether or not certain variables can be used to predict selection to colonel, this thesis predicts selection to colonel based on metrics created by "conventional wisdom." These metrics are discussed in the data description in Chapter III.

A description of the layout of the remaining chapters in this thesis follows. Chapter II provides a literature review. The focus of the literature review is on the application of logistic regression with emphasis placed on its use to predict selection for advancement in military applications. Chapter III is used to describe the data utilized in this study. The focus of this chapter is on the composition of each observation and highlights the summary statistics associated with variables in the study. Chapter IV provides the description and results of the data analysis performed for the thesis. This chapter defines the logistic regression process and introduces the systematic development and fit of models for this study. The three best fit models are highlighted and explained. The thesis concludes with Chapter V, which provides a summary of results and identifies the potential for future studies.

[2] As witnessed by this researcher's 25 years of uniformed service, taken from historical common knowledge, and highlighted by Kizilkaya (2004).

THIS PAGE INTENTIONALLY LEFT BLANK

# II. LITERATURE REVIEW

## A. INTRODUCTION

Logistic regression is a powerful data analysis tool for modeling outcomes of a Binomial random variable. Thus, logistic regression is an effective tool for modeling successes versus failures in a variety of applications. Promotion is an example of a success versus failure response variable. Promotion can be modeled as a Bernoulli random variable where 1 corresponds to the event an individual is selected for promotion and 0 corresponds to the event an individual is not selected for promotion. In this chapter, we identify studies that use logistic regression to model response variables with a binary response. In addition to discussing several examples found in the literature, we also identify published works that use logistic regression to study what variables influence an individual's chance for promotion in a military ranking system.

## B. LOGISTIC REGRESSION

Logistic regression models are found in a great variety of fields. The following three examples illustrate the use of logistic regression in three separate areas: medical outcome prediction, sociological status modeling, and athletic performance analysis.

Rush (2001) studies the factors influencing retinopathy of prematurity, a disease associated with blindness primarily found in premature infants and is the binary response variable for the study. The factors analyzed in this study numbered 29 and were discrete or categorical in nature. The use of logistic regression aided in identifying the risk factors closely associated to this disease, thus allowing medical practitioners to properly assess patients' conditions. Rush's model further debunked a factor formerly considered one of the critical risk factors. Similar to the study in Rush (2001), the analysis in this thesis aims to determine if critical factors associated with the AGR can be used to predict selection to colonel.

Another example of logistic regression is found in Achia, Wangombe, and Khadioli (2010). They assess the factors associated with sociologic status. They use logistic regression to examine the determining factors of poverty in Kenya. The study

digs deeper than the three indicators commonly thought to categorize poverty and assess a variety of additional variables. Principal components analysis is used to reduce the number of variables in this study. The resulting logistic regression model is derived from six variables, all showing significance in their influence on determining the poverty probability. The results of Achia, Wangombe, and Khadioli (2010) highlights the importance of augmenting factors that capture "common wisdom" associated with economic status identification with other factors.

Clark, Johnson, and Stimpson (2013) study the conventional wisdom behind football field goal successes. The 11 variables considered in the field goal study provide the basis for Clark, Johnson, and Stimpson's model. Their model both discredits conventional wisdom and provides a method to better predict field goal classifications. Their use of logistic regression for outcome predictions and conventional wisdom validation is similar in methodology, as seen in Chapter IV of this thesis.

In addition to the three studies described above, examples of the use of logistic regression in a military application are also prevalent in the literature. Two examples provided here are the applications of logistic regression to career decisions after the Naval Academy and military retention modeling.

As external pressures continue to weigh heavy on individuals in the military, the choice to stay in the military is of interest to the force structure managers. Turner (1990) examined the factors leading to a nurse's choice. Faced with an increased demand for nurses coupled with a reduction in enrollments to the program, Turner investigates the critical influences necessary to narrow the gap. Fifteen variables are used to fit a logistic regression model which predicts with 98.7% accuracy, a nurse's choice to stay or leave. Further, the logistic regression gives only a 1.2% False-Positive rate and a 1.7% False-Negative rate. Yet, even with these results, Turner suggests the addition of more focused variables to potentially aid in developing improved retention tools. Turner's use of a confusion matrix to compute False-Positive and False-Negative rates is used in this study, and is found in Chapter IV.

Burroughs (2007) explored the influences behind a Naval Academy Midshipman's selection of service in the Marine Corps as opposed to becoming a submariner. Burroughs developed 10 categories to derive the independent variables when considering service selection. His final model had eight independent variables. The results of a binary logistic regression identified a clear delineation between the influences factoring in to a midshipmen's selection for service. The logistic regression accurately predicted 79.85% of the selections for the Marine Corps and 85.1% for those selecting the subsurface community. Burroughs admits his study was narrow in focus and should be broadened to include additional variables. His use of logistic regression to identify criteria influential to the leadership selection process is similar to the methodology studied in this thesis.

## C. PROMOTION

Logistic regression models are useful, as exemplified by the previous documents, to identify critical influencers, to predict studied events, and to validate standard practices. In this section, four documents are highlighted for their use of logistic regression in aspects related to military promotions. These examples provide insight into the techniques and methodologies conducted in this thesis.

The earliest opportunities for promotion or advancement experienced by military officers are found at the Academy's, Senior Reserve Officer Training Corps programs and/or enlistment. Fox (2003) considers the midshipmen leadership selections of the United States Naval Academy. The main focus of Fox's work is to assess how well selections for the brigade midshipmen leadership are met. By means of qualitative research and analysis, Fox identified three general categories utilized in leadership selection. A logistic regression model comprised of eight variables created from the three general categories determined the selection of brigade midshipmen leadership as meeting the desired end state. That is to say, midshipmen leadership is being selected based upon intended expectations of a leader. This technique, to validate common practices, is similar to the conventional wisdom validation found in Chapter IV of this thesis. Fox also

concluded there may be more than just the eight variables involved in leadership selection (2003).

Kizilkaya (2004) addresses the relationship between commissioning sources and the retention to the grade of O-4, major, and promotion to the grades of O-4 and O-5, lieutenant colonel. Focusing specifically on the promotion models, five general categorical variables are chosen to generate the two logistic regression models. Variables are screened based upon relevancy to the study, data accuracy, and data field voids. Kizilkaya uses nine variables in his models and their adequacy is measured by means of goodness-of-fit and misclassification rates. The final models achieve contradictory results when comparing the O-4 and O-5 promotion models. Even though the sources of commissioning are identified as determining factors for promotion, the contrasting outcomes raise more questions than answers.

A more recent study of promotion model predictions is found in Gonzalez’s (2011) lieutenant colonel promotion and command selection rates. Gonzalez utilizes a logistic regression model with 32 of variables to produce the fitted models supporting his findings. The models’ accuracy is validated by means of the resulting $R^2$ values and misclassification rates. The three models generated produced at best an accuracy of 87% selection to lieutenant colonel. Gonzalez’s findings identify significant variables and whether or not serving in combat is relative to promotion selection. Like Gonzalez, this thesis uses the misclassification rate as a critical part of a model’s measure of performance.

Weko and Pontius (2012) examined the criteria necessary for selection to colonel. Their work considered the relevant factors influencing the selection process of packets submitted by Army Active Guard/Reserve lieutenant colonels. As did Fox (2003) in assessing midshipmen leadership, Weko and Pontius aligned the relevant factors associated in colonel selection to that of the conventional wisdom of the time (2012). Weko and Pontius (2012) found no combination of factors guarantees colonel selection; however, they did attribute one factor to possessing the most influence in selecting colonels. They examined 21 variables: five of which are identified as representing conventional wisdom. Six of the 21 variables were deemed to be the most influential.

Three of the six align themselves with conventional wisdom, while one of those is not an actual conventional wisdom variable, but is used to derive it (Weko & Pontius, 2012).

## D. SUMMARY

Logistic regression models are useful, in the identification of critical influencers, the accurate prediction of studied events, and the validation of standard practices. The study conducted by Weko and Pontius (2012) is the inspiration for and provides the backdrop to this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

# III. DATA

## A. INTRODUCTION

The data used for the analysis in this thesis is provided by PA&E. The data is compiled from 1144 individual packets of lieutenant colonels submitted for promotion to colonel across three selection boards between FY10 and FY12. The 1144 promotion packets correspond to 684 individuals according to the identification number included in the data. If a packet went before more than one board it is indicative of that packet having not been selected during the previous board. That packet may or may not have been selected in the subsequent board. All duplicate packets are deleted, leaving only the most recently considered packet. The data contains 59 input variables. In this study, only 33 of the variables are used. The omitted fields are either duplicates of existing fields or contain information irrelevant to this study.

The Naval Postgraduate Schools Human Research Protection Program requires an Institutional Review Board (IRB) examine all studies conducted involving individuals and/or information related to an individual. The resulting IRB used in this study determined the data contained no personal identification information. Additionally, individual records are identified by an anonymous identification number, thus the study is exempt to the full IRB protocol.

The identification number coupled with the board number and board year are used to reduce the 1144 packets to one packet for each of 684 separate individuals having submitted packets for selection review. The 684 individuals correspond to 321 one-time submissions, 266 two-time board submissions, and 97 three-time board submissions (reference Table 1). A total of 170 packets were selected for promotion to colonel; representing 25% of all individual packets submitted as selected over the three-year period. The variable, Selected, is a binary variable indicating whether or not an individual's packet was selected, "1," or was not selected, "0." This is the categorical response variable for the purpose of this study.

Table 1. Frequency of Selection Packet Submissions–depicts the total number of packets by the number of times an individual packet went before the selection board. The table further identifies the selection percentage according to the number of times a packet is submitted.

| Times Submitted | Total Packets | Selected | |
|---|---|---|---|
| | | Yes | No |
| 1 | 321 | 29% | 71% |
| 2 | 266 | 24% | 76% |
| 3 | 97 | 12% | 88% |
| TOTAL | 684 | 25% | 75% |

The board identification number is composed of three distinct numbers and is only used in identifying the board-year each packet was considered for and whether a file was reviewed in one, two or all three of the selection boards.

## B. VARIABLES

In this section, we discuss the independent variables in the data analysis. The logistic regression models are used to determine if any of these variables provide the ability to predict whether or not a submitted package results in a promotion.

The variable labeled Education is a binary variable identifying whether an individual is educationally qualified, "1," or non-educationally qualified, "0." For an individual to be educationally qualified, they must have completed all required military courses for their branch and/or career field. Six-hundred-fifty-four of the 684 packets submitted were academically qualified.

The variable Zone accounts for a packet's zone of consideration. A packet is either above the zone, in the primary zone, or below the zone. For this categorical variable an above the zone is represented by a "1," a primary zone is represented by "0," and a below the zone is represented by a "-1." For a packet to be considered below the zone the packet is reviewed during the 3- to 4-year time-in-grade time period as a lieutenant colonel. The primary zone of consideration is typically within the five-year mark time-in-grade as a lieutenant colonel and is considered as the normal look time for

selection for promotion. For a packet to be considered above the zone, the packet is reviewed beyond the five-year time-in-grade mark as a lieutenant colonel. The number of packets considered below the zone is 171, as seen in Table 2. The number of packets considered within the primary zone is 225. The number of packets considered above the zone is 288.

Table 2. Zones of Consideration–depicts the total number of packets submitted by consideration zone and the selection rate percentage.

| Zone | Total Packets | Selected | |
|---|---|---|---|
| | | Yes | No |
| Above | 288 | 20% | 80% |
| Primary | 225 | 47% | 53% |
| Below | 171 | 4% | 96% |

The variable Gender is a binary variable where "1" represents male and "0" represents female. Females account for 128 or 18.7% of the packets submitted for selection, as seen in Table 3, with 29 being selected. Males account for the remaining 556 or 81.3% of the packets with 141 being selected.

Table 3. Gender–Identifies the number of packets by sex and compares them to the number of packets selected within the each category. Female (F); Male (M).

| | # | Selected | Not Selected |
|---|---|---|---|
| **F** | **128** | 23% | 77% |
| **M** | **556** | 25% | 75% |

Age is a numeric variable accounting for the age of the individual upon submission of the packet to the selection board. Figure 3 illustrates the distribution of the age groups considered in this study.

AGE
(YEARS)

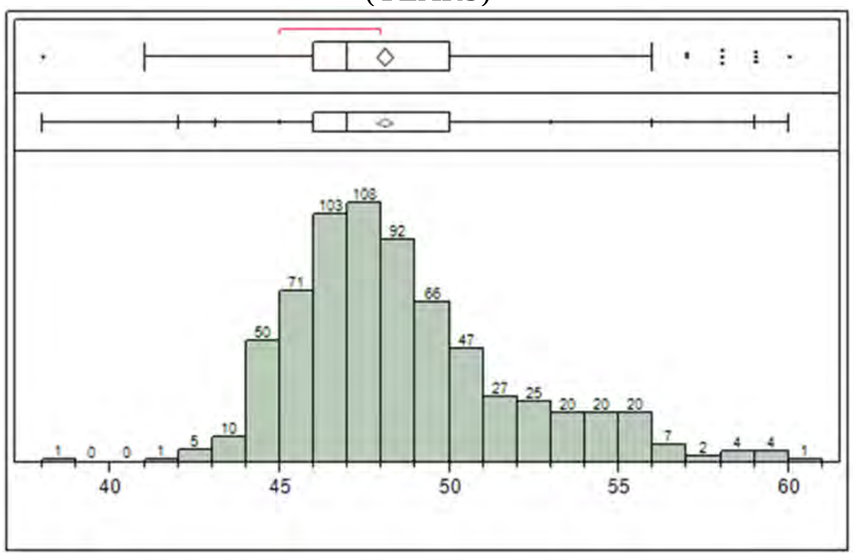


Figure 3. Age–depicts the number of individual packets by the reported age at the time the packet was submitted. The data is graphically represented in an Outlier and Standard Quartile Box-Plot as well as a Histogram. The box-plots identify the average age as 48.07 $\pm$ 3.26 years. Thirty-eight outliers exist above the age of 55 and one at age 38. The histogram reflects what appears to be a normal distribution with a positive skew in the results.

The Time-in-Service variable identifies the length of time an individual has served in the military at the time of the packets submission and its distribution is depicted in Figure 4.

TIME IN SERVICE
(Years)

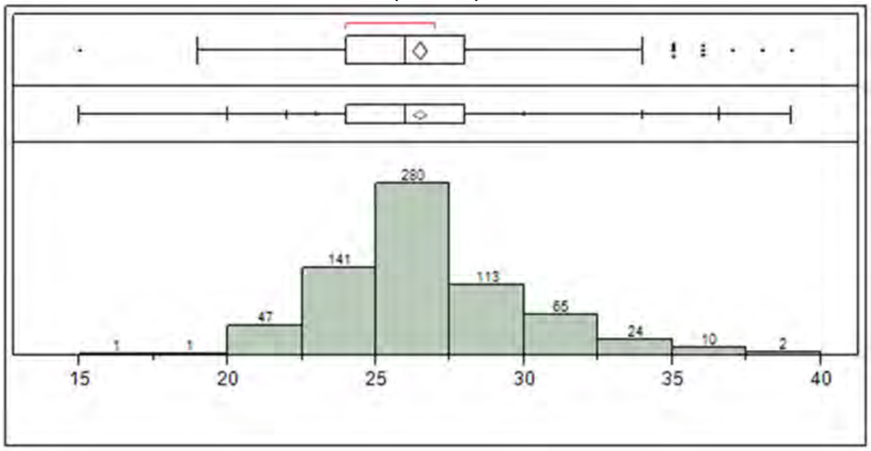


Figure 4. Time in Service–as measured in years, depicts the number of individual packets relative to the total years of military service. The data is graphically represented in an Outlier and Standard Quartile Box-Plot as well as a Histogram. The box-plots identify the average time-in-service as 26.46 ± 3.12 years with 50% of the packets representing 24 to 28 years of service. Several outliers exist at 35 years and beyond, as well as one outlier at 15 years. The histogram reflects near-normal results.

The Tape variable is a binary representation of whether or not an individual required a body-fat composition measurement or "taping" as it is commonly referred to. Zero represents no requirement for a taping and accounts for 310 of the packets submitted. One indicates that an individual required taping and accounts for 374 of the packets. Tape is derived from a formula accounting to an individual's height and weight based on standardized tables. If an individual's weight exceeds the maximum required weight according to a height index, the individual is then "taped," where a sequential series of body dimensions are measured and calculated to determine the individual's body-fat composition. Those not meeting the standards are placed on a program to correct the problem and are denied special recognition (i.e., awards, special training, and promotions). Of those requiring taping 79 are selected for promotion. Of those not requiring taping 91 are selected.

The Security Clearance variable is a binary variable of whether an individual possesses a Top Secret level clearance. Individuals possessing a Top Secret clearance are

represented by a “1” and account for 431 of the packets submitted, of which 134 are selected. Of the remaining 253 not possessing a Top Secret clearance, 36 are selected.

The variable Airborne accounts for those individuals having completed airborne training and earning the right to wear the parachutist badge. To be Airborne qualified, an individual must complete five (5) successful parachute jumps from an aircraft at an altitude of not less than 1000 feet at the culmination of a three-week training period. This variable was converted from a categorical yes or no to a binary “1” or “0,” respectively. Of the 366 airborne qualified individuals 104 are selected for promotion, whereas only 66 of the remaining 318 non-airborne qualified individuals are selected.

The variable Awards>Meritorious Service Medal (MSM) is a binary variable where “1” accounts for 325 of the packets having at least one award greater than an MSM, 111 having been selected. Zero represents the remaining 359 packets with at least one MSM or lower award, with 59 having been selected.

The number of Deployments Post-2001 is a variable representing the number of deployments within a range of 0 to 5 years for each packet submitted. Figure 5 and Table 4 depict the number of packets submitted according to the number of deployments conducted since 2001. One-hundred twenty-nine of the 402 individuals deployed were selected for promotion. Seventy-six percent of those selected were deployed.

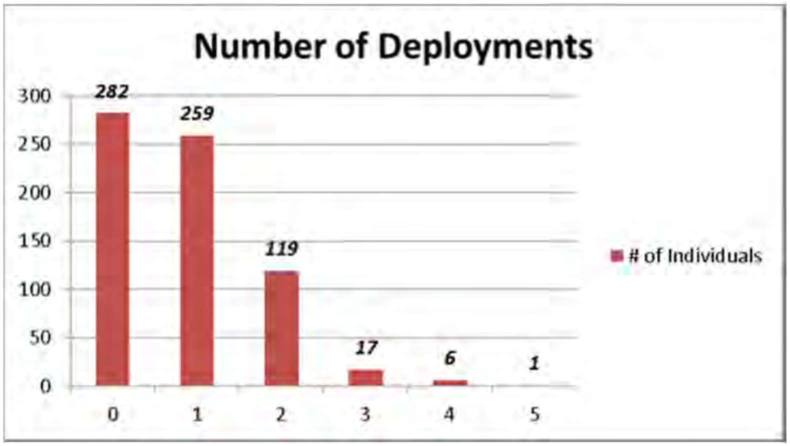


Figure 5. Deployments Post-2001–depicts the number of packets submitted according to the number of deployments conducted since 2001.

Table 4. Selection Rate for Deployments Post-2001–depicts the percentage rate of the number of packets submitted according to the number of deployments conducted since 2001.

| Deployed x Times | Total Packets | Selected | |
|---|---|---|---|
| | | Yes | No |
| 0 | 282 | 15% | 85% |
| 1 | 259 | 30% | 70% |
| 2 | 119 | 37% | 63% |
| 3 | 17 | 35% | 65% |
| 4 | 6 | 17% | 83% |
| 5 | 1 | 0% | 100% |

Longest Deployment variable represents the greatest length of time, in consecutive months, an individual is deployed. The deployments range from 0 to 17 months. The average deployment length is 5.74 $\pm$ 5.33 months. The strong majority, 73.4% of the packets submitted, either did not deploy (41.2%) or deployed for more than 11 months (32.2%).

Senior Service College (SSC) is a binary representation of whether or not an individual completed the next level of military education required to attain the rank of a flag officer. Forty-six of the 76 having completed SSC are selected for promotion (reference Figure 6). The graph divides the data into its separate senior service colleges: the National War College (NWC); the Army War College (AWC); College of Naval Warfare (CNW); Senior Service College Fellowship (SSC F); Joint Advanced Warfighter Course (JAWS); Industrial College of the Armed Forces (ICAF); Army War College Distance-Learning (AWC DL).

SENIOR SERVICE COLLEGE (SSC) COMPLETION

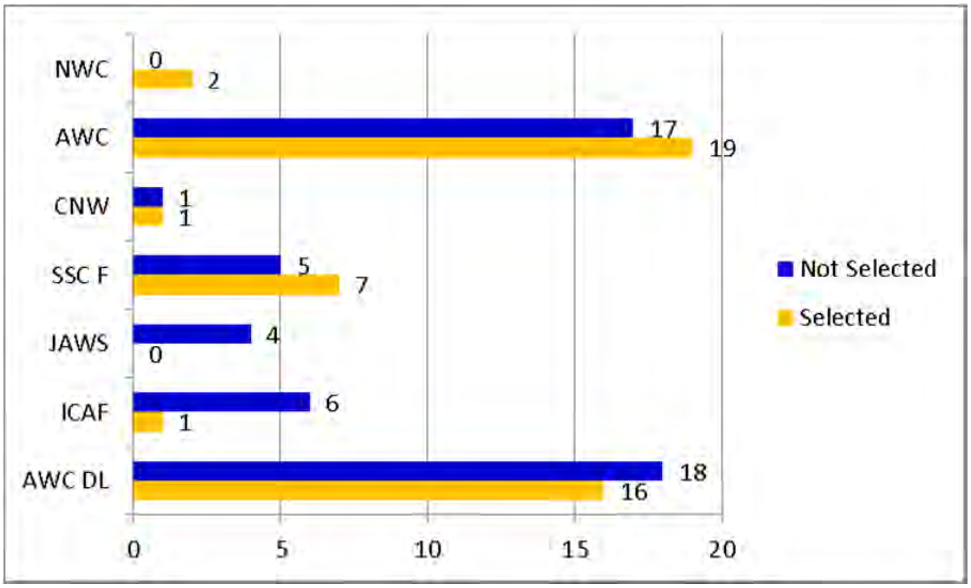


Figure 6. Senior Service College (SSC) Completion–depicts the number of packets submitted having completed SSC. The graph compares the total number Selected (represented in Gold) to the number Not Selected (represented in Blue). These totals are distributed across the various Senior Service Colleges.

The master's variable is a binary variable to identify whether or not an individual has completed a master's degree. Those having completed a master's are represented by a "1" and account for 430 of the packets, 134 of which are selected. Thirty-six of the remaining 254 not having a master's degree are selected for promotion.

The variable Battalion Command is a binary variable indicating those packets having at least one battalion command as a lieutenant colonel, as accounted for by a "1." One-hundred-thirteen individuals had battalion command of which 58 are selected. One-hundred-twelve of the 571 packets not having battalion command are selected for promotion.

The variable Lieutenant Colonel Ratings accounts for the total number of ratings an individual received while at the grade of lieutenant colonel. This variable is used as a baseline to establish percentages for the remaining variables capturing various rating statistics.

The Percentage of General Officers Ratings is derived from the total number of ratings received by a lieutenant colonel from a general officer or the civilian equivalent of a flag officer and the total number of lieutenant colonel ratings overall.

The Percentage of General Officer Above Center of Mass Ratings is derived from the total number of general officers ratings categorized above center of mass for that lieutenant colonel and the total number of lieutenant colonel ratings overall.

The Percentage of Deployed Above Center of Mass Ratings is derived from the total number of ratings categorized above center of mass while deployed as a lieutenant colonel and the total number of lieutenant colonel ratings overall.

Percent Total Above Center of Mass is derived from the total number of ratings lieutenant colonel received in the category above center of mass and the total number of lieutenant colonel ratings overall.

Longest Time-on-Station (ToS) is a variable that represents the longest total number of consecutive months an individual remained within the boundaries of one duty station. The data for this variable falls within the range of 0 to 161 months with an average monthly ToS of $47.15 \pm 23.38$ months. Thirty-seven individuals report a ToS of 90 months or greater.

The categorical variable labeled Married, referenced below in Table 5 and Figure 7, identifies whether an individual, at the time of each packet's submission, is Married (M); Divorced (D); Single (S); Widowed (W).

Table 5. Marital Status–Identifies the number of packets by Marital Status and compares them to the number of packets selected within the each group. Married (M); Divorced (D); Single (S); Widowed (W).

| | # | Selected | Not Selected |
|---|---|---|---|
| M | 548 | 26% | 74% |
| D | 67 | 18% | 82% |
| S | 67 | 19% | 81% |
| W | 2 | 0% | 100% |

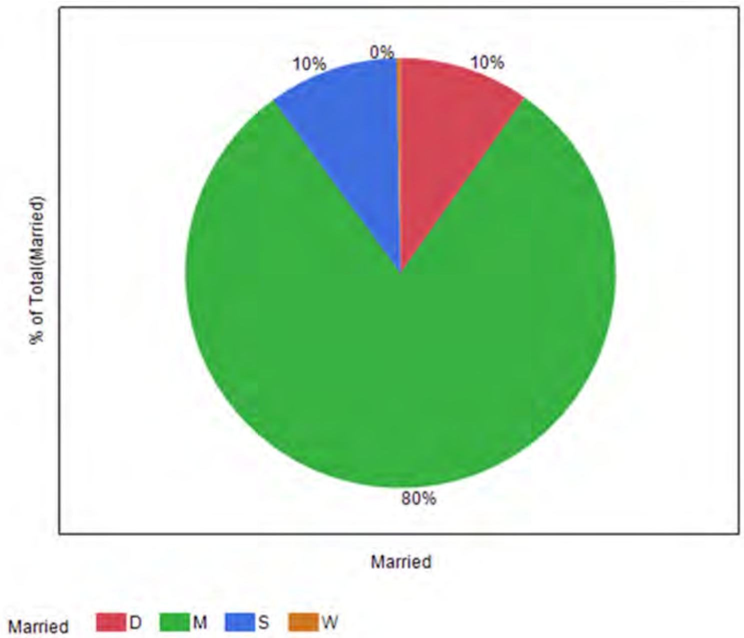


Figure 7. Graphically depicts the Marital Status breakdown of the packets submitted by Married (M); Divorced (D); Single (S); Widowed (W).

The categorical variable labeled Race, as seen in Table 6 and Figure 8 identifies whether an individual is ethnically affiliated as White (W); Black (B); Hispanic (H); Filipino (F); Asian (A); Native American (N); or Pacific Islander (P).

Table 6. Race–Identifies the number of packets by ethnicity and compares them to the percentage of packets selected within the ethnic group. White (W); Black (B); Hispanic (H); Filipino (F); Asian (A); Native American (N); or Pacific Islander (P).

| | # | Selected | Not Selected |
|---|---|---|---|
| W | 459 | 28% | 72% |
| B | 158 | 16% | 84% |
| H | 45 | 20% | 80% |
| A | 9 | 33% | 67% |
| P | 9 | 33% | 67% |
| F | 2 | 0% | 100% |
| N | 2 | 0% | 100% |

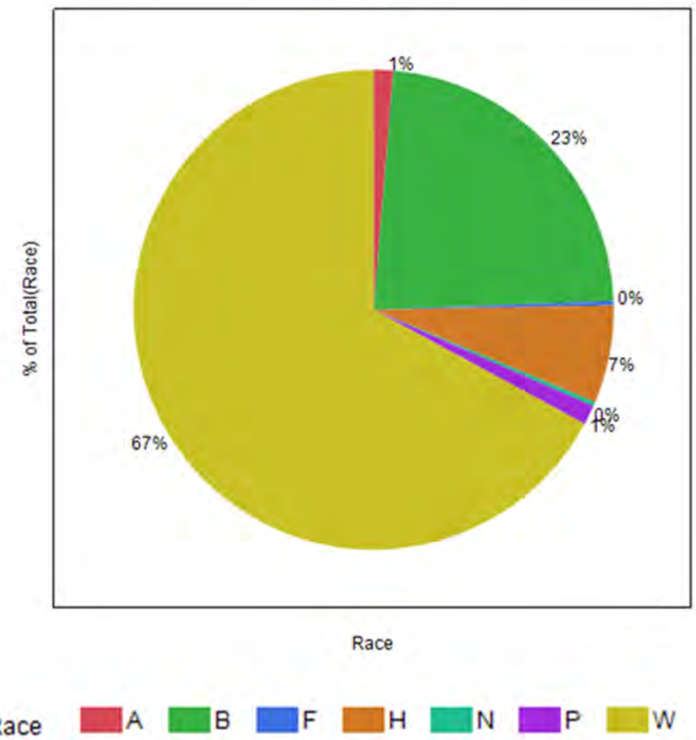


Figure 8. Graphically depicts the ethnic breakdown of the packets submitted by White (W); Black (B); Hispanic (H); Filipino (F); Asian (A); Native American (N); or Pacific Islander (P).

The categorical variable labeled Branch identifies the regimental affiliation an individual has based upon their military training. Table 7 identifies each of the regimental affiliations within the data set.

Table 7. Branch–tabulates the individual Regimental Affiliations against the number of packets whether or not they were selected.

| | # | Selected | Not Selected |
|---|---|---|---|
| Logistics (LG) | 195 | 27% | 73% |
| Adjutant (AG) | 77 | 16% | 84% |
| Engineers (EN) | 68 | 25% | 75% |
| Civil Affairs (CA) | 62 | 37% | 63% |
| Signal (SC) | 48 | 21% | 79% |
| Military Intelligence (MI) | 46 | 17% | 83% |
| Infantry (IN) | 32 | 25% | 75% |
| Aviation (AV) | 28 | 46% | 54% |
| Finance (FI) | 24 | 25% | 75% |
| Military Police (MP) | 24 | 25% | 75% |
| Field Artillery (FA) | 23 | 17% | 83% |
| Chemical (CM) | 19 | 5% | 95% |
| Armor (AR) | 14 | 21% | 79% |
| Psychological Operations (PO) | 7 | 43% | 57% |
| Air Defense Artillery (AD) | 5 | 20% | 80% |
| Quartermaster (QM) | 5 | 20% | 80% |
| Dental (DC) | 2 | 0% | 100% |
| Transportation (TC) | 2 | 0% | 100% |
| Medical Service (MS) | 1 | 0% | 100% |
| Ordnance (OD) | 1 | 100% | 0% |
| Special Forces (SF) | 1 | 0% | 100% |

## C. CONVENTIONAL WISDOM

Conventional Wisdom is an additional collection of six variables added to the original data set and includes what is perceived to be, at the time this data set was developed, to be the five most needed traits in order to be selected for promotion to colonel. These variables are derived from a compilation of five of the previously

described variables. Five of the newly derived variables are all a binary variables where "1" accounts for the possession of the variable trait and "0" its opposite.

The first in this new set of variables is Conventional Wisdom 1 (CW1), this is the completion of SSC and is a straightforward conversion from the SSC binary representation. The second is Conventional Wisdom 2 (CW2) and accounts for whether or not an individual was deployed. This is derived from the longest deployed variable and translates any numeric value greater than zero to the binary representation for being deployed, "1." The third is Convention Wisdom 3 (CW3) and is a straightforward binary translation for completion of a master's degree. The Fourth is Conventional Wisdom 4 (CW4) and again is a straightforward binary translation from the battalion command, accounting for whether or not an individual was in a command position as a lieutenant colonel. The fifth variable is Conventional Wisdom 5 (CW5), and accounts for whether or not an individual possesses ACOM ratings greater than 75%. This variable is a "1" if the percent total above center of mass value is greater than or equal to 75%. The final variable added to the conventional wisdom set is the Percent Total Conventional Wisdom (%CW). This variable assesses an individual's overall percentage of possession of the conventional wisdom variables and is represented as a numeric variable.

As depicted in Tables 8 and 9, only four individuals possess all the criteria necessary to be labeled as having met conventional wisdom. Of the 680 not meeting all the criteria for conventional wisdom, 166 are selected for promotion.

Table 8. Conventional Wisdom–tabulates the individual Conventional Wisdom criteria and identifies the number having been Selected or not Selected according to whether meeting Conventional Wisdom or not.

| | Met | | | Not Met | | |
|---|---|---|---|---|---|---|
| | # | Selected | Not Selected | # | Selected | Not Selected |
| **CW1** | **76** | 61% | 65% | **608** | 20% | 390% |
| **CW2** | **402** | 32% | 212% | **282** | 15% | 588% |
| **CW3** | **430** | 31% | 221% | **254** | 14% | 606% |
| **CW4** | **113** | 51% | 95% | **571** | 20% | 410% |
| **CW5** | **130** | 65% | 55% | **554** | 16% | 544% |
| **ALL** | **4** | 100% | 0% | **680** | 24% | 310% |

Table 9. Conventional Wisdom vs. Selected–compares the numbers of packets having met all criteria to be classified as Conventional Wisdom to the number of packets having been selected.

| | | Selected | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| Conventional Wisdom | Yes | 4 | 0 | 0.6% |
| | No | 166 | 514 | 99.4% |
| | Total | 24.9% | 75.1% | 684 |

# IV. ANALYSIS/RESULTS

## A. INTRODUCTION

We use logistic regression (Hosmer, Lemeshow, & Sturdivant, 2013) models to estimate the probability of selection to colonel as a function of selection criteria and their two-factor interactions. In these models the binary response variable, Selected, is modeled as $Y_1$, $Y_2$,…, $Y_{684}$ independent Bernoulli variables with respective probabilities of promotion $P_1$, $P_2$,…, $P_{684}$. Logistic regression models link these probabilities to the dependent variables with the logistic link function

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \ ,$$

where, here, the subscripts indicating individual observations are suppressed, $x_1$, $x_2$,…, $x_k$ are the $k$ dependent variables, (which may include numeric variables, categorical variables and interactions) and $\beta_0$, $\beta_1$,…, $\beta_\kappa$ are the parameters to be estimated. The inverse logit function is used to express the probabilities as a function of the dependent variables.

$$P = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_k x_k)}}$$

Thirty of the 33 variables identified in Chapter III are used for the purpose of fitting models, while the three remaining are used solely to distinguish between the three different selection board years and each individual submission. Table 10 describes the selection criteria variables used throughout this study in the fitting process and identifies the variables by their modeling type.

Table 10. Selection Criteria Variable Description and Type.

| Variable | Description | Type |
|---|---|---|
| $Y$ | Categorical Response - Selected | Nominal |
| $X_{ED}$ | Military Education Qualified | Nominal |
| $X_{ZONE}$ | Zone of Consideration | Nominal |
| $X_{GEN}$ | Gender | Nominal |
| $X_{AGE}$ | Age | Numeric |
| $X_{TIS}$ | Time-in-Service | Numeric |
| $X_{TAPE}$ | Tape Required | Nominal |
| $X_{SC}$ | Security Clearance | Nominal |
| $X_{ABN}$ | Airborne Qualified | Nominal |
| $X_{MSM}$ | Award > Meritorious Service Medal | Nominal |
| $X_{\#DEPL}$ | # of Deployments Post-2001 | Numeric |
| $X_{LD}$ | Longest Deployment | Numeric |
| $X_{SSC}$ | Senior Service College | Nominal |
| $X_{MSTR}$ | Master's Degree Completed | Nominal |
| $X_{BN}$ | Battalion Command | Nominal |
| $X_{RATE}$ | # of Lieutenant Colonel Ratings | Numeric |
| $X_{\%GO}$ | Percent General Officer (GO) Ratings | Numeric |
| $X_{\%GA}$ | Percent GO Above-Center-of-Mass (ACOM) Ratings | Numeric |
| $X_{\%DA}$ | Percent Deployed ACOM Ratings | Numeric |
| $X_{\%TA}$ | Percent Total ACOM Ratings | Numeric |
| $X_{TOS}$ | Longest Time-on-Station | Numeric |
| $X_{MAR}$ | Marital Status | Nominal |
| $X_{RACE}$ | Race | Nominal |
| $X_{BR}$ | Branch (Military Specialty) | Nominal |
| $X_{CW1}$ | Conventional Wisdom 1 | Nominal |
| $X_{CW2}$ | Conventional Wisdom 2 | Nominal |
| $X_{CW3}$ | Conventional Wisdom 3 | Nominal |
| $X_{CW4}$ | Conventional Wisdom 4 | Nominal |
| $X_{CW5}$ | Conventional Wisdom 5 | Nominal |
| $X_{\%CW}$ | Percent Total Conventional Wisdom | Numeric |

Thirteen models are fit, each based on a different initial set of dependent variables as described in this chapter. Backwards elimination is used to eliminate unneeded or redundant predictor variables, with the criteria that variables with p-values less than 0.1 are retained. The resulting thirteen models fit are then assessed based on misclassification rates, as described in the next section.

## B. MEASURES OF EFFECTIVENESS

Misclassification rates are computed by means of a confusion matrix, a table used to compute performance measures for comparing predicted outcomes to the actual recorded results. The confusion matrix is based on the probabilities of selection for each individual in the data set estimated from the logistic regression fit. Individuals whose estimated probabilities of selection are above a threshold value are classified (predicted) as being selected for promotion. Table 11 is an example confusion matrix taken from the analysis of Model 1 (in the Appendix). The accurately predicted results are highlighted in green and for the purpose of this study are classified as being Correct, based on a 0.5 threshold. The 483 predicted to not be selected are accurately identified, along with the 110 predicted to be selected are actually selected and comprise the classification of Correct. Those predicted to be selected, the 31 highlighted in yellow, but are actually not selected are classified as False-Positive. The remaining 60, highlighted in tan, are predicted as not to be selected yet were actually selected and are classified as being False-Negative.

Table 11. Confusion Matrix example taken from the results generated from Model 1 in the Appendix.

| | | Predicted | |
|---|---|---|---|
| | | No | Yes |
| Actual | No | 483 | 31 |
| | Yes | 60 | 110 |

The three measures of effectiveness used in this study focus on the prediction percentages associated with being Correct, False-Positive and False-Negative. The minimum acceptable False-Positive rate is 1% and the minimum acceptable False-Negative rate is 15%. The combination of the False-Positive and False-Negative outcomes is used to identify the ideal threshold of the confusion matrix for each fitted model. The correct prediction percentage is used to compare fitted model outcomes.

We use five thresholds—0.5, 0.6, 0.7, 0.8, 0.9—for predicting a selection board outcome. The threshold is manually adjusted to analyze the results for 0.5 to 0.9 thresholds inclusively. An Excel spreadsheet is used to tabulate the 0.5–0.9 threshold confusion matrices. A sample of the spreadsheet is seen here, Figure 9, depicting the actual promotion selection results under the selected column as a Yes/No response. The estimated probability of selection is in decimal form, as seen next to the “Prob.Sele” of Figure 9. The final four columns in Figure 9 show the predicted outcome based on thresholds 0.6, 0.7, 0.8, 0.9. For each threshold, a confusion matrix is computed to visually determine at which threshold value the acceptable False-Positive and False-Negative percentages occur.

| Selected | Prob. Sele | Thresh=.5 | Thresh=.6 | Thresh=.7 | Thresh=.8 | Thresh=.9 |
|---|---|---|---|---|---|---|
| No | 0.185768 | No | No | No | No | No |
| Yes | 0.873004 | Yes | Yes | Yes | Yes | No |
| No | 0.000725 | No | No | No | No | No |
| Yes | 0.873004 | Yes | Yes | Yes | Yes | No |
| No | 0.185768 | No | No | No | No | No |
| Yes | 0.648481 | Yes | Yes | No | No | No |
| No | 0.098251 | No | No | No | No | No |
| No | 0.015015 | No | No | No | No | No |
| Yes | 0.381448 | No | No | No | No | No |
| No | 0.008848 | No | No | No | No | No |
| Yes | 0.72456 | Yes | Yes | Yes | No | No |
| Yes | 0.150997 | No | No | No | No | No |
| Yes | 0.690934 | Yes | Yes | No | No | No |
| Yes | 0.771402 | Yes | Yes | Yes | No | No |
| No | 0.002227 | No | No | No | No | No |
| No | 0.018887 | No | No | No | No | No |
| No | 0.055887 | No | No | No | No | No |
| No | 0.001511 | No | No | No | No | No |
| Yes | 0.892823 | Yes | Yes | Yes | Yes | No |
| Yes | 0.134472 | No | No | No | No | No |
| Yes | 0.969795 | Yes | Yes | Yes | Yes | Yes |

Figure 9. Sample Excel Spreadsheet taken from Model 1 used to create threshold confusion matrices

For example, in Figure 9, the arrows highlight a board-selected packet with a predicted probability of selection of 38%, clearly not achieving the threshold of 0.5 (50%) or higher, thus it will not be classified as a predicted select. Yet, the packet

highlighted by the stars possesses a 69% predicted selection probability, obviously greater than both 50 and 60% but not 70% and above. This predicted selection is then classified as selected for only the 0.5 and 0.6 thresholds.

## C. MODELS

Thirteen models are systematically fit, from the list of independent variables, with the goal of identifying the criteria necessary for promotion selection and determining if conventional wisdom is viable in selection prediction. Each model is processed by means of the SAS Institute Incorporated, JMP® Pro 10.0.0 64-bit Edition. All 13 models and their analysis are found in the Appendix: Model Development.

The best-fit models are chosen based on their measures of effectiveness in comparison to the remaining models. These models are the top performers based on their possession of the fewest variables necessary among those which have acceptable thresholds for one or more threshold-levels and for an 85% or greater percentage Correct.

Ten of the 13 models have 85% accuracy. For two models, all five threshold levels yield greater than 85% accuracy. Four models contain four, one model contains three, and three models contain two threshold levels with an accuracy of 85% or greater. When comparing models based upon the number of acceptable classification rates, two models possesses four or more; two possessed two; and three possessed one acceptable classification rate.

Using the binary variable—identifying whether a packet was selected or not—as the response variable, the models below are constructed from a selection of the 30 predictor variables established in Table 10. Model A, derived from 6B in the Appendix, contains 15 of the original variables and 57 two-factor interactions. Model B, derived from Model 6, contains eight of the original variables. Model C, derived from Model 3, contains only the five Conventional Wisdom variables.

### 1. MODEL A

The first of these models uses all the original selection variables and their two-factor interactions. Backwards elimination gives a final model with 15 of the original

variables and 57 two-factor interactions. Model A's Misclassification Rate is 0.0307 with all thresholds having acceptable values, as highlighted in Table 12. Of significance, the 0.9 threshold has a 0% False-Positive rate.

Table 12. Threshold Comparison-Model A.

| MODEL A | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 96.93% | 97.37% | 96.35% | 95.32% | 94.30% |
| % False Pos | 1.46% | 0.58% | 0.29% | 0.15% | 0.00% |
| % False Neg | 1.61% | 2.05% | 3.36% | 4.53% | 5.70% |

### 2. MODEL B

The second of these models takes into account all the original variables only. After backwards elimination, only eight of the original variables remain, as seen in Table 13, Parameter Estimates. The Misclassification Rate for the final model is 0.1072 and with an acceptable threshold of 0.8 (Table 14).

Table 13. Parameter Estimates for Model B with corresponding standard errors (Std Error), likelihood ratio test statistics (ChiSquare) for the inclusion of the parameter, and the p-value (Prob>ChiSq) for the test.

**Parameter Estimates**

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept | -0.46468758 | 2.8378876 | 0.03 | 0.8699 |
| Zn #[-1] | -3.22662287 | 0.4093103 | 62.14 | <.0001* |
| Zn #[0] | 2.0546601 | 0.2501092 | 67.49 | <.0001* |
| Age | -0.17352285 | 0.0587183 | 8.73 | 0.0031* |
| Long DEP | 0.1441119 | 0.0293344 | 24.13 | <.0001* |
| SSC[0] | -0.8512225 | 0.2063814 | 17.01 | <.0001* |
| MSTR[0] | -0.51642608 | 0.1693845 | 9.3 | 0.0023* |
| BN CMD[0] | -0.51063159 | 0.17745 | 8.28 | 0.0040* |
| LTC Ratings | 0.3107868 | 0.0957486 | 10.54 | 0.0012* |
| % Total ACOM | 7.8327551 | 0.8320895 | 88.61 | <.0001* |

Using the parameter estimates from Table 13, the fitted final model takes the form:

$$\hat{y} = -0.4647 - 3.227x_{ZONE[-1]} + 2.055x_{ZONE[0]} - 0.1735x_{AGE} + 0.1441x_{LD} - 0.8512x_{SSC[0]}$$

$$-0.5164x_{MSTR[0]} - 0.5106x_{BN[0]} + 0.3108x_{RATE} + 7.833x_{\%TA} \quad ,$$

where ŷ is the estimates log odds of the probability of selection, and the independent variables, the x's, are identified by their subscripts.

The estimated log-odds can then be used to compute the estimated probability of selection. The three level categorical variable zone is represented by two binary variables, $x_{ZONE[-1]}$ which is 1 if zone = -1 (below zone) and 0 otherwise and $x_{ZONE[0]}$ which is 1 if zone = 0 (in the primary zone) and 0 otherwise. For example, a packet submitted with the criteria: In the Primary Zone – 0; Age – 45; Longest Deployment – 17; not completed SSC – 0; has a Master's – 1; not have Battalion Command – 0; LTC Ratings – 6; % Total ACOM Ratings – 0.83 gives an estimated probability of 97.7%. Since the Zone variable is represented by a "0", $x_{ZONE[-1]}$ = 0 and $x_{ZONE[0]}$ = 1:

$$\hat{y} = -0.4647 - 3.227(0) + 2.055(1) - 0.1735(45) + 0.1441(17) - 0.8512(1)$$

$$-0.5164(-1) - 0.5106(1) + 0.3108(6) + 7.833(0.83)$$

$$= 3.75088$$

and to compute the estimated probability, $\hat{P}$

$$\hat{P} = \frac{1}{1+e^{-\hat{y}}} = \frac{1}{1+e^{-3.75088}} = 0.97704 \quad .$$

Based on the confusion matrix comparison thresholds, this example is correctly predicted for all thresholds.

Table 14. Threshold Comparison-Model B.

| MODEL B | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 89.33% | 89.18% | 88.16% | 86.70% | 83.04% |
| % False Pos | 4.53% | 3.51% | 2.05% | 0.88% | 0.58% |
| % False Neg | 6.14% | 7.31% | 9.80% | 12.43% | 16.37% |

In this example the numeric variables of longest deployment, lieutenant colonel ratings, and percent total above center of mass increase the probability of selection as the variable increases in value. The numeric variable age decreases the probability of selection as the value increases. The binary variables of senior service college, master's, and battalion command all increase the probability of selection when the packet is in possession of either of the variables. Adjusting the zone of consideration results in an increase when in the primary zone and a decrease if in the other zones.

### 3. MODEL C

The third model looks at the only the Conventional Wisdom variables for their influence on promotion selection. Backwards elimination yields the model with five variables and resulting in parameter estimates listed in, Figure 10. The Misclassification Rate for this final model is 0.1813 and did not possess an acceptable threshold. Model C is examined based on transforming the associated original variable to a binary Yes "1" / No "0" value.

**Parameter Estimates**

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept | -0.0381279 | 0.1907037 | 0.04 | 0.8415 |
| CW1[0] | -0.8269004 | 0.1512865 | 29.87 | <.0001* |
| CW2[0] | -0.4927269 | 0.1176953 | 17.53 | <.0001* |
| CW3[0] | -0.3653776 | 0.1256494 | 8.46 | 0.0036* |
| CW4[0] | -0.4438859 | 0.1291372 | 11.82 | 0.0006* |
| CW5[0] | -1.056559 | 0.1209755 | 76.28 | <.0001* |

Figure 10. Parameter Estimates for Model C Conventional Wisdom Variables and with corresponding standard errors (Std Error), likelihood ratio test statistics (ChiSquare) for the inclusion of the parameter, and the p-value (Prob>ChiSq) for the test.

These variables produce a final model taking the form:

$$\hat{y} = -0.0381 - 0.8269x_{CW1[0]} - 0.4927x_{\mathrm{CW2[0]}} - 0.3654x_{\mathrm{CW3[0]}} - 0.4439x_{\mathrm{CW4[0]}} - 1.057x_{\mathrm{CW5[0]}} \ ,$$

This equation can now be applied to the data. Taking an example from the data, a packet submitted with the criteria: CW1 Yes – 1; CW2 Yes – 1; CW3 Yes – 1; CW4 No – 0; CW5 No – 0, produces a 53.7% estimated probability of selection.

For Model C, CW1 = 1 corresponds to $x_{\mathrm{CW[0]}}$ = -1 and CW1 = 0 corresponds to $x_{\mathrm{CW[0]}}$ = 1. The same applies to each of the CW variables. Therefore substituting example packet variables in to (1) yields:

$$\hat{y} = -0.0381 - 0.8269(-1) - 0.4927(-1) - 0.3654(-1) - 0.4439(1) - 1.057(1)$$

$$= 0.146$$

to compute the estimated probability, $\hat{P}$

$$\hat{P} = \frac{1}{1+e^{-0.146}} = 0.536435 \quad .$$

Based on the confusion matrix, this model has unacceptable False-Negative rates for all thresholds and unacceptable False-Positive rates for all but the 0.9 threshold.

Table 15. Threshold Comparison-Model C.

| MODEL C | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 81.87% | 80.56% | 78.36% | 77.78% | 76.46% |
| % False Pos | 5.56% | 4.24% | 2.34% | 1.61% | 0.44% |
| % False Neg | 12.57% | 15.20% | 19.30% | 20.61% | 23.10% |

The combination of variables in this model has influence on the probability of selection. Assessing the variables individually, suggests the possession of only a single binary variable trait favors Percent Total Above Center of Mass with a 24.8% probability of selection and is found in 84 of the 170 selected packets. The remaining variables' probabilities of selection (for the individuals possessing only that respective trait) are: Senior Service College at 17.2% as found in 46 packets, Longest Deployment at 9.6% as found in 129 packets, Battalion Command at 8.8% as found in 58 packets, and Master's at 7.6% as found in 134.

An individual possessing all variable traits has an estimated probability of selection at 95.9%, while a model possessing no traits has an estimated probability of selection of 3.8%. The number of packets with all five traits numbered four out of the 170 selected for promotion and the packets with no traits numbered two.

### 4. MODEL D

The final model is fitted with only the Percent Total Conventional Wisdom variable. For this model the misclassification rate is at 0.1901 and once again no

acceptable threshold comparison is observed, Table 17. This model demonstrates a 0% False-Positive, for the 0.9 threshold.

Table 16. Parameter Estimates for Model D with corresponding standard errors (Std Error), likelihood ratio test statistics (ChiSquare) for the inclusion of the parameter, and the p-value (Prob>ChiSq) for the test.

**Parameter Estimates**

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept | -3.6250715 | 2.837888 | 0.03 | <.0001* |
| %CW | 0.0635459 | 0.83209 | 88.61 | <.0001* |

Using the parameter estimates from Table 16, the final model takes the form:

$$\hat{y} = -3.625 + 0.0635x_{\%CW} \quad ,$$

Taking an example from the data, a packet submitted having met three of the five CW criteria or 60% CW; the model is re-written as follows,

$$\begin{aligned} \hat{y} &= -3.625 + 0.0635(60) \\ &= 0.188 \end{aligned}$$

giving, $\hat{P}$

$$\hat{P} = \frac{1}{1+e^{-0.188}} = 0.546862 \quad .$$

Based on the confusion matrix comparison thresholds, this example is correctly predicted for the 0.5 threshold and incorrectly predicted, as a False-Negative for the remaining thresholds.

Table 17. Threshold Comparison-Model D

| MODEL D | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 80.99% | 77.78% | 77.78% | 77.78% | 75.73% |
| % False Pos | 7.60% | 2.19% | 2.19% | 2.19% | 0.00% |
| % False Neg | 11.40% | 20.03% | 20.03% | 20.03% | 24.27% |

When examining the Conventional Wisdom traits, using this model, an individual has an estimated probability of selection ranging from 2.6% to 93.9%. A packet submitted with no Conventional Wisdom traits registers a 2.6% probability of selection. Transitioning from zero to one Conventional Wisdom trait increases the selection probability to 8.7%. As a packet increases to all five Conventional Wisdom traits, the probability raises to 25.3% for two traits, 54.7% for three, 81.1% at four, and finally a 93.9% probability of selection with all five Conventional Wisdom traits.

## D. SUMMARY

Logistic regression analysis is used to fit 13 models where the response variable is selection for promotion to colonel. The models are generated from a mixed composition of single and two-factor interactions of 29 independent variables. The models are processed by means of automated and manual backwards elimination. Four of the 13 models are presented in the analysis section and their effectiveness is assessed.

Acceptable classification rates are established based upon a percent Correct value of at least 85%, a False-Positive of 1% or less and False-Negative of 15% or less. Two of the four models examined in this chapter meet this target and Model B is the better of the two models. Model A is not considered since the model is over fit with 72 independent variables. Thus it is discarded, even though it met the target for all five thresholds and possessed over 90% accuracy in all threshold levels.

Model B's findings suggest one's zone of consideration, age, longest deployment, senior service college completion, possession of a master's degree, battalion command, number of ratings as a lieutenant colonel, and the total percentage above center of mass

ratings have a significant influence on selection. The results demonstrate an accuracy of prediction ranging from 83.04% to 89.33% with a False-Positive rate of 0.58% to 4.53%.

Model C's findings suggest all conventional wisdom variables, whether or not an individual possess the trait, influences the prediction for selection. The accuracy of prediction ranges from 76.46% to 81.87% with a False-Positive rate of 0.44% to 5.56% and a False-Negative rate of 12.57% to 23.10%. Model C comes close to being replicated in its results by those of Model D, which only accounts for the Percent Total Conventional Wisdom. The results of these conventional wisdom models are not as significant as Model B, based on the acceptable classification rates. It is perceived that only individuals possessing all conventional wisdom traits are subject for selection, however, the results of this study would suggest otherwise.

THIS PAGE INTENTIONALLY LEFT BLANK

# V. CONCLUSION/FUTURE WORK

## A. CONCLUSION

This thesis provides data analysis on the selection process of the FY 2009–2011 Army Active Guard/Reserve (AGR) colonel selection boards and determines conventional wisdom's role in the process. Logistic regression analysis is conducted on the 684 individual packets submitted to three consecutive selection boards. A single logistic regression model is identified with the capability of predicting selection with 86.7% accuracy.

The results of this study concur with Weko and Pontius' (2012) original finding that "Relevant factors conformed with Conventional Wisdom." All five of the original selection criteria associated with Conventional Wisdom are relevant to the selection process and contained in Model B. While not a guarantee, the results of this thesis do suggest promotion selection is predictable to 83.04–89.33% accuracy and presents at a False-Positive rate of at worst 4.53% versus Tse's 16% (1993).

Weko and Pontius further stated the most important factor associated with AGR colonel selection is an individual's performance ratings. This study suggests to the contrary. Even though nine of the 13 models contain some degree of promotion ratings, these findings are not significant enough to suggest performance rating as being the most important factor. Four of the 29 independent variables considered in the models can be attributed to performance rating. Only three models contain three of the four attributed performance rating variables. When considering the best-fit model, only one of the attributed performance rating variables made it into the eight-variable fitted model. If we are to consider the over-fit top model in this study, at best, 34.72% of the significant variables were associated in one fashion or another with performance rating.

Conventional Wisdom plays a role in the selection process. When considered solely on its own, conventional wisdom's influence on selection is predicted, at best, with 82% accuracy while incorrectly predicting a selection up to 7%.

## B. FUTURE WORK

The conclusions of this thesis concur with Weko and Pontius (2012). The data reviewed by Weko and Pontius and analyzed in this thesis examined only one skill badge, the parachutist badge (Airborne). There are over 20 skill badges at various levels within their categories. Consideration could also be given to the variety of other decorations, awards and honors.

A closer look should be given to the Officer Evaluation Report (OER). Per conversations with Weko and Pontius, some of the OERs rated ACOM are assessing the officer for less than 12 months. The identification of referred reports and any other derogatory paperwork would/should have an impact on selection. Additionally, taking into account the number of ratings received by a single rater along with the number of positions held by the rated officer, may present an influencing factor to promotion. Accounting for deployment as a lieutenant colonel and the OERs associated may also present themselves as influencers.

Another consideration is to take into account the needs of the field. That is to say, what quotas account for the positions requiring to be filled? Quotas by demographics, whether branch affiliation, gender, race, or skill identifiers. Also, what are the current demands for the Army as a whole and how do they affect the Army Reserve and thus the AGR system. Are there draw-downs, do budget cuts have an effect?

# APPENDIX. MODEL DEVELOPMENT

| MODEL 1 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 86.70% | 85.38% | 84.06% | 82.16% | 79.39% |
| % False Pos | 4.53% | 3.07% | 1.75% | 1.02% | 0.44% |
| % False Neg | 8.77% | 11.55% | 14.18% | 16.81% | 20.18% |

***Threshold Comparison-Model 1 begins with nine main effects from the original selection criteria and the newly added percent conventional wisdom variable. The nine main effects were selected based upon their summary statistics' observations. Backwards elimination yields the final resulting model comprised of three of the original selection criteria and the percent conventional wisdom variable. The Misclassification Rate for this final model is 0. 1330 and did not possess an acceptable threshold comparison target value intersection.***

| MODEL 1A | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 86.11% | 85.53% | 84.65% | 82.60% | 79.39% |
| % False Pos | 4.53% | 3.07% | 1.90% | 1.17% | 0.44% |
| % False Neg | 9.36% | 11.40% | 13.45% | 16.23% | 20.18% |

***Threshold Comparison-Model 1A takes the resulting model from Model 1 above and adds in the two-factor interactions. Backwards elimination yields the final resulting model comprised of two of the original selection criteria, the percent conventional wisdom variable and a single two-factor interaction. The Misclassification Rate for the final model is 0.1389 and did not possess an acceptable threshold comparison target value intersection.***

| MODEL 2 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 86.40% | 85.67% | 84.80% | 82.60% | 78.80% |
| % False Pos | 5.85% | 4.09% | 1.61% | 1.17% | 0.58% |
| % False Neg | 7.75% | 10.23% | 13.60% | 16.23% | 20.61% |

***Threshold Comparison-Model 2 revisits the original Model 1 and added to it the two-factor interactions. Backwards elimination yields the final resulting model comprised of significant p-values for two of the original selection criteria, the percent conventional wisdom variable and 3 two- factor interactions. The Misclassification Rate for the final model is 0.1360 and did not possess an acceptable threshold comparison target value intersection.***

| MODEL 3 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 81.87% | 80.56% | 78.36% | 77.78% | 76.46% |
| % False Pos | 5.56% | 4.24% | 2.34% | 1.61% | 0.44% |
| % False Neg | 12.57% | 15.20% | 19.30% | 20.61% | 23.10% |

***Threshold Comparison-Model 3 is processed with only the newly generated five Conventional wisdom variables. These variables all possessed significant p-values and have a Misclassification Rate of 0.1813. However, as with the previous models, did not possess an intersection of the acceptable threshold comparison target values.***

| MODEL 4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 82.02% | 80.99% | 79.68% | 76.46% | 75.73% |
| % False Pos | 7.02% | 5.85% | 2.05% | 0.58% | 0.15% |
| % False Neg | 10.96% | 13.16% | 18.27% | 22.95% | 24.12% |

***Threshold Comparison-Model 4 expanded on model three and added the Conventional wisdom variables' two-factor interactions. Backwards elimination yields the final resulting model comprised of all five conventional wisdom variables and two of their two-factor interactions. The Misclassification Rate for this model is 0.1798 and as with its predecessor, did not possess an intersection of the acceptable threshold comparison target values.***

| MODEL 5 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 80.99% | 77.78% | 77.78% | 77.78% | 75.73% |
| % False Pos | 7.60% | 2.19% | 2.19% | 2.19% | 0.00% |
| % False Neg | 11.40% | 20.03% | 20.03% | 20.03% | 24.27% |

***Threshold Comparison-Model 5 fits a model with only the percent conventional wisdom variable. Once processed a Misclassification Rate is at 0.1901 and once again no intersection of threshold comparison target values is observed. However, of significance, this is the first model to show a 0% false positive, as seen at the 0.9 threshold.***

| MODEL 6 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 89.33% | 89.18% | 88.16% | 86.70% | 83.04% |
| % False Pos | 4.53% | 3.51% | 2.05% | 0.88% | 0.58% |
| % False Neg | 6.14% | 7.31% | 9.80% | 12.43% | 16.37% |

***Threshold Comparison-Model 6 analyzes only the original selection criteria; it did not take into account the newly generated conventional wisdom criteria. Backwards elimination yields the final resulting model comprised eight of the original selection criteria and a 0.1072 Misclassification Rate. The model a possessed acceptable threshold comparison target value intersection at the 0.8 threshold.***

| MODEL 6A | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 90.20% | 89.91% | 88.60% | 86.40% | 83.48% |
| % False Pos | 4.24% | 3.07% | 2.49% | 1.46% | 0.58% |
| % False Neg | 5.56% | 7.02% | 8.92% | 12.13% | 15.94% |

***Threshold Comparison-Model 6A. The final results for Model 6 were then used along with their two-factor interactions generate Model 6A. Backwards elimination yields the final resulting model comprised six of the original criteria from Model 6 and adds 5 two-factor interactions. Model 6A's Misclassification Rate is a 0.0984 with suggested acceptable thresholds of 0.8, according to this studies threshold comparison target values.***

| MODEL 6B | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 96.93% | 97.37% | 96.35% | 95.32% | 94.30% |
| % False Pos | 1.46% | 0.58% | 0.29% | 0.15% | 0.00% |
| % False Neg | 1.61% | 2.05% | 3.36% | 4.53% | 5.70% |

***Threshold Comparison- Model 6B is derived from all the original selection criteria and their two-factor interactions. Backwards elimination yields the final resulting model comprised 15 of the original criteria and 57 two-factor interactions. Model 6B's Misclassification Rate is 0.0308 with all thresholds possessing the threshold comparison target values. Of great significance, the 0.9 threshold possesses a 0% false positive. Even though this value is shared with Model 5, Model 6B is 18.63% more accurate in the percent correct category.***

| MODEL 7 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 94.74% | 94.74% | 94.30% | 92.69% | 90.79% |
| % False Pos | 2.34% | 1.75% | 1.02% | 0.29% | 0.29% |
| % False Neg | 2.92% | 3.51% | 4.68% | 7.02% | 8.92% |

***Threshold Comparison- Model 7 takes into account all the original selection criteria, the Conventional wisdom variables and all two-factor interactions. Backwards elimination yields the final resulting model comprised 11 of the original selection criteria and 43 of their two-factor interactions. The Misclassification Rate for the final model is 0.0529 and possessed acceptable threshold comparison target values between the 0.6 and 0.9 thresholds inclusively.***

| MODEL 8 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 88.16% | 86.99% | 86.26% | 84.65% | 82.31% |
| % False Pos | 4.68% | 3.51% | 1.90% | 1.17% | 0.44% |
| % False Neg | 7.16% | 9.50% | 11.84% | 14.18% | 17.25% |

***Threshold Comparison- Model 8 is comprised of the original selection criteria with the Conventional Wisdom variables and is absent of the original selection criteria associated with each individual Conventional Wisdom variable. Backwards elimination yields the final resulting model comprised 7 of the original selection criteria and all five Conventional Wisdom variables. The Misclassification Rate for the final model is 0.1189 and possessed acceptable threshold comparison target value intersection at the 0.7 threshold.***

| MODEL 8A | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 89.47% | 89.77% | 88.60% | 86.99% | 84.21% |
| % False Pos | 4.24% | 3.22% | 1.90% | 1.17% | 0.44% |
| % False Neg | 6.29% | 7.02% | 9.50% | 11.84% | 15.35% |

***Threshold Comparison- Model 8A takes the results of Model 8 above and all of its two-factor interactions. Backwards elimination yields the final resulting model comprised 3 of the original selection criteria, three conventional wisdom variables and 15 two-factor interactions. The Misclassification Rate for the final model is 0.1057 and possessed acceptable threshold comparison target value intersection at the 0.7 & 0.8 thresholds.***

| MODEL 8B | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|
| % Correct | 90.64% | 90.79% | 89.62% | 87.13% | 84.06% |
| % False Pos | 3.80% | 2.63% | 1.75% | 1.32% | 0.88% |
| % False Neg | 5.56% | 6.58% | 8.63% | 11.55% | 15.06% |

***Threshold Comparison- Model 8B looks at the original starting conditions for Model 8 and adds their two-factor interactions. Backwards elimination yields the final resulting model comprised three of the original selection criteria, three conventional wisdom variables and 25 of their two-factor interactions. The Misclassification Rate for this final model is 0.0940 and possessed an acceptable threshold comparison target value intersection at the 0.7 & 0.8 thresholds.***

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF REFERENCES


Achia, T. N. O., Wangombe, A., & Khadioli, N. (2010). A logistic regression model to identify key determinants of poverty using demographic and health survey data. *European Journal of Social Sciences*, 13(1), 38–45.

Burroughs, J. C. (2007). *An exploratory study of persuasive influences on midshipman service selection at the United States Military Academy: Comparing service selection for the submarine force and the United States Marine Corps* (Master's thesis, Naval Postgraduate School). Retrieved from http://calhoun.nps.edu/public/handle/10945/3484

Clark, T. K., Johnson, A. W., & Stimpson, A. J. (2013). Going for three: Predicting the likelihood of field goal success with logistic regression*.* Presented at the 7th Annual MIT Sloan Sports Analytics Conference. Cambridge, MA.

England, T. F. (1984). The Active Guard/Reserve program: A new military personnel status. *Military Law Review,* 6, 1–76

Fox, J. P. (2003). *Selecting the brigade leadership at the United States Naval Academy: Who are the Stripers?* (Master's thesis, Naval Postgraduate School). Retrieved from http://calhoun.nps.edu/public/handle/10945/985

Gonzalez, M. D. (2011). *Analysis of promotion rates to lieutenant colonel and selection for command for USMC aviation supply and maintenance officers* (Master's thesis, Naval Postgraduate School). Retrieved from http://calhoun.nps.edu/public/handle/10945/10613

Headquarters Department of the Army (2004). *The Active Guard Reserve (AGR) Program, Department of the Army* (Army Regulation 135–18). Washington, DC: Author.

Headquarters Department of the Army (2010). *Commissioned Officer Professional Development and Career Management* (Department of the Army Pamphlet 600–3). Washington, DC: Author.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: Wiley.

Kizilkaya, Z. (2004). *An analysis of the effect of commissioning sources on retention and promotion of U.S. Army officers* (Master's thesis, Naval Postgraduate School). Retrieved from http://calhoun.nps.edu/public/handle/10945/1560

Rush, S. W. (2001). *Logistic regression: The standard method of analysis in medical research*. Retrieved from http://ramanujan.math.trinity.edu/tumath/research/studpapers/s3.pdf

Tse, L. T. (1993). *Application of logistic regression and survival analysis to the study of CEP, manpower performance and attrition* (Master's thesis, Naval Postgraduate School). Retrieved from http://calhoun.nps.edu/public/handle/10945/39968

Turner, P. B. (1990). *Retention in the Navy Nurse Corps* (Master's thesis, Naval Postgraduate School). Retrieved from http://calhoun.nps.edu/public/handle/10945/27694

United States General Accounting Office (1982). *Information on military technician conversions to Full-Time Active Duty Guard And Reserve*. Retrieved from http://www.gao.gov/assets/140/138635.pdf

Weko, C., & Pontius, J. (2012, July). Analysis of Fiscal Year 2009, 2010, and 2011 Active Guard/Reserve colonel boards. Presented at the Office of the Chief, Army Reserve. Washington, DC.

# INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California